FIRST ESA WORKSHOP ON COMPUTER VISION AND IMAGE PROCESSING FOR SPACEBORNE APPLICATIONS

ABSTRACT BOOK

10-12 June 1991
ESTEC, Noordwijk
The Netherlands

SESSION 5A : MODEL BASED VISION
CHAIRMAN : P. Plancke

Locating modelled 3D objects from Monocular Perspective Brightness Images for Autonomous Navigation or Tracking.
M. Dhome, J.T. Lapreste and all - Univ. of Clermond-Ferrand, France

Methodologies and Techniques for Interpretation of 3-D Range Images.
F. Solina, Z. Bjelogrlic - Intecs Sistemi, Italy

Tracking with a Robot Head.
G. Casalino, G. Germano, G. Sandini - DIST-LIRA, Genova, Italy

Navigation Updating of Autonomous Air Vehicles using 3D Object Recognition.
O. Reichert, R. Goulette - Sagem, France

The VIEWS Project: its Relationship to Spaceborne Applications.
V. Stanger - GEC, UK

UTOPIA, a Software Tool for Parallel Processing Applications.
L. Thieling, W. Ameling - CWA, Germany

Space Robotic Control Requirements and Abilities Identification Methodology.
Y. Cruvellier - Fabricom

METHODOLOGIES AND TECHNIQES FOR INTERPRETATION OF 3D RANGE IMAGES

Franc Solina[1] and Zaviša Bjelogrlić

INTECS Sistemi s.p.a
Via Vertumno 2c
00157 ROMA, Italy

ABSTRACT

We present two techniques of interpretation of range images that might apply to rover and sampling missions using techniques of model based vision. Applications range from navigation and obstacle avoidance to grasping of natural objects. The first technique is modeling of objects using volumetric models, the second is segmentation using surface parametric patches.

Keywords: shape recovery, segmentation, parametric shape models

1 INTRODUCTION

One of the major goals of computer vision is to recover descriptions of the physical world that enable locating, handling and identifying objects. Since shape information plays a crucial part in these activities, a substantial effort has been devoted to identify proper models for shape representation. Different shape reconstruction methods introduced different shape models, that is models that fit into the particular reconstruction philosophy (bottom-up, top-down or a combination of the two). For a detailed overview of different reconstruction methods see [8].

Segmentation of images into regions corresponding to single objects or their parts is one of the harder problems in computer vision. Recognition of objects would be easier for a vision system if the system knew which areas in an image correspond to single objects. Segmentation, on the other hand, would also be simpler if the identity of objects in the scene and hence their shape could be found beforehand. It is not obvious which problem should be tackled first. Model-based object recognition systems using feature indexing, which have the advantage of knowing the exact models of objects in the scene, try to identify these objects on the basis of some very specific features first. These local features or combinations of them can be used either to instantiate an object model from a data base or for further aggregation into shape models of larger granularity. However, this does not apply to unconstrained or unknown environment where we have to assume that no apriori known object models are given aside from generic models that encompass a large set of all possible part shapes—a vocabulary for describing the scene. We believe that in such cases the solution for segmentation might well be to do it *simultaneously*—to recover such parts in the images that can be described with a selected part shape vocabulary [2].

Most approaches to segmentation in computer vision are based on using local image information, in the form of low level image models such as edges, surface patches and surface normals. Segmentation methods can be divided into boundary and region based methods. Boundary methods try to find significant changes that separate regions in images, while region based methods look for similarity which indicates elements that belong together. When 3-D data is available, surface normals or surface discontinuities (C_0 and C_1) are a commonly used local features. Partitioning then involves thresholding using histogram analysis or clustering in multidimensional space when several properties are used simultaneously. Since this features are very local, noise and missing information makes these segmentation methods unreliable. The problem can be partially alleviated by using coherence measures in a somewhat larger neighborhood. Examples are edge tracking and region growing and using consistency criteria for merging and splitting. Fitting of planar or higher order surface patches in a local neighborhood is a popular method to assure local consistency in range maps. Some of these methods derive the initial boundaries of local surface patches

[1] Also at: University of Ljubljana, Computer Vision Laboratory, Faculty of Electrical Engineering and Computer Science, Tržaška 25, 61001 Ljubljana, Slovenia

from edges and significant changes in surfaces expressed in terms of differential geometry or discontinuities of surface depth and surface normals. The resulting segmentation is often *arbitrary*, even if the similar neighboring surface patches are later merged, which is especially true for nonpolyhedral objects. This is because merging or growing of such small surface regions essentially still relies on local information. If such local segmentation methods are made sensitive enough to detect subtle changes in first or second derivatives in order to find part boundaries, they become susceptible to noise and details that are not relevant for the targeted level of representation.

Much work has been done recently on the problem of reconstructing piecewise-smooth surfaces in one or more dimensions [7,15] which is posed as an optimization problem. In all these approaches the data is weighted uniformly which means that the algorithms do not possess the capabilities to adapt to different conditions in different parts of the image. The global measure, provided by the energy function, is not able to tell which parts of the image are well described in terms of the underlying models and which are not. Also it is difficult to see how these approaches could be extended to subsequent stages of the vision problem without using models with fewer degrees of freedom.

All local segmentation methods, used so far in computer vision, based whether on surfaces or on contours, have problems with arbitrary segmentation. The reason seems to be that an essentially local piece of information cannot decide on the shape of the whole part if the concept of the whole part as such is not defined.

The problem of using part boundaries to define the shape of parts can be circumvented by directly defining a family of all possible part shapes. Biederman [6] argued that human perception uses a set of primitive building blocks which can describe the wealth of different shapes by combining them like phonemes in a language. Perceptual grouping is another way to extract the relevant low level image features and filter out the noise in order to reduce the search space when model matching is performed [13]. But the predictive power of generic models is not used to its full potential when only rules for combining low level models into larger ones are used. If the higher level generic models are well defined, one can attempt to find them in a more direct way. Search can be made more efficient if the objective can be defined in purely mathematical terms. In this paper we focus on two particular approaches (and possible extensions) to range image interpretation that we are involved in

1. using recovery of superquadric models [17] a technique that has already found tentative applications in sorting of mail pieces [16], in the Mars Rover project [10] and in grasping with robot hands in general [1],

2. segmentation as the search for the best description in terms of primitives [11]..

2 VOLUMETRIC MODELS

Superquadrics are an extension of basic quadric surfaces and solids [3]. Superquadric surface is defined by the following equation

$$F(x,y,z) = \left(\left(\left(\frac{x}{a_1} \right)^{\frac{2}{\epsilon_2}} + \left(\frac{y}{a_2} \right)^{\frac{2}{\epsilon_2}} \right)^{\frac{\epsilon_2}{\epsilon_1}} + \left(\frac{z}{a_3} \right)^{\frac{2}{\epsilon_1}} \right)^{\epsilon_1}. \tag{1}$$

When both ϵ_1 and ϵ_2 are 1, the surface defined is an ellipsoid. When $\epsilon_1 \ll 1$ and $\epsilon_2 = 1$, the superquadric surface is shaped like a cylinder. Parallelepipeds are produced when both $\epsilon_1 \ll 1$ and $\epsilon_2 \ll 1$. Modeling capabilities of superquadrics can be enhanced by deforming them in different ways, such as tapering and bending [4]. Some examples of superquadric models are in Figure 1.

The function in equation (1) is called the inside-outside function because it determines where a given point (x,y,z) lies relative to the superquadric surface. If $F(x,y,z) = 1$, point (x,y,z) lies on the surface of the superquadric. If $F(x,y,z) > 1$, the corresponding point lies outside and if $F(x,y,z) < 1$, the corresponding point lies inside the superquadric.

The inside-outside function (1) defines the superquadric surface in an object centered coordinate system (x_S, y_S, z_S). Since 3-D points in range images are expressed in an image coordinate system, an inside-outside function for general position requires additional six parameters to account for position and orientation:

$$F(x,y,z) = F(x,y,z; a_1, a_2, a_3, \epsilon_1, \epsilon_2, \phi, \theta, \psi, p_x, p_y, p_z). \tag{2}$$

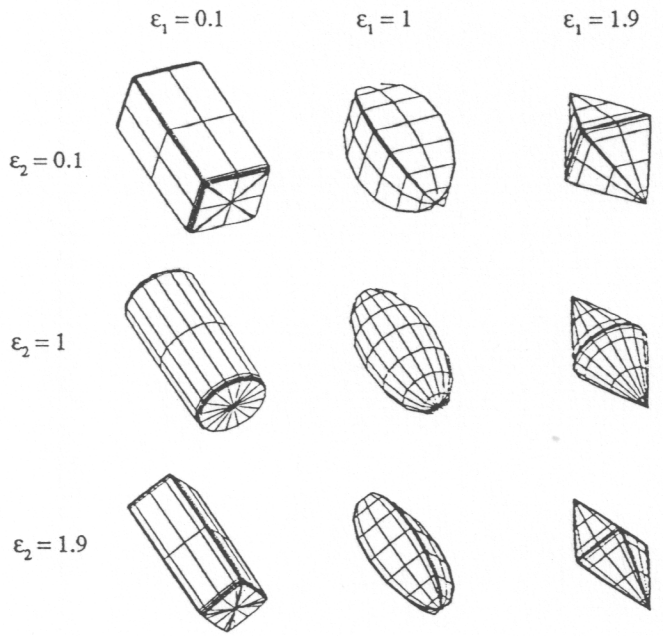

Figure 1: *Shape of superquadric models as a function of shape parameters*

This expanded inside-outside function has 11 parameters; a_1, a_2, a_3 define the superquadric size; ϵ_1 and ϵ_2 are shape parameters; ϕ, θ, ψ define the orientation in space, and p_x, p_y, p_z define the position in space. We refer to the set of all model parameters as $\Lambda = \{a_1, a_2, \ldots, a_{11}\}$.

2.1 Recovery of superquadrics

For shape recovery assume at the moment that some simple segmentation is performed so that all given range points lie on the surface of the object to be modeled. Suppose we have N 3-D surface points (x_W, y_W, z_W) which we want to model with a superquadric. We want to vary the 11 parameters $a_j, j = 1, \ldots, 11$ in equation 2 to get such values for a_j's that most of the 3-D points will lie on, or close to the model's surface. There will probably not exist a set of parameters Λ that perfectly fits the data. Finding the model Λ for which the distance from points to the model's surface is minimal is a least-squares minimization problem. Since due to self occlusion, not all sides of an object are visible at the same time, we have to introduce an additional constraint. Among all possible solutions we want to find the *smallest* superquadric that fits the given range points in the least squares sense. We define the following function which has a minimum corresponding to the smallest superquadric that fits a set of 3-D points *and* a function value for surface points which is known before minimization

$$R = \sqrt{a_1 a_2 a_3}\,(F - 1), \tag{3}$$

Since, for a point (x_W, y_W, z_W) on the surface of a superquadric

$$R(x_W, y_W, z_W; a_1, \ldots, a_{11}) = 0, \tag{4}$$

we have to find

$$G = min \sum_{i=1}^{N} [R(x_{W_i}, y_{W_i}, z_{W_i}; a_1, \ldots, a_{11})]^2. \tag{5}$$

Since R is a nonlinear function of 11 parameters $a_j, j = 1, \ldots, 11$, minimization must proceed iteratively. Given a trial set of values of model parameters Λ_k, we evaluate equation (3) for all N points and employ a procedure to improve the trial solution. The procedure is then repeated with a set of new trial values Λ_{k+1} until the sum of least squares (5) stops decreasing, or the changes are statistically meaningless. In most cases 15 iterations are more then sufficient[2].

[2]We use the Levenberg-Marquardt method for nonlinear least squares minimization since first derivatives $\delta R/\delta a_i$ for $i = 1, \ldots, 11$ can be computed analytically.

Only very rough initial estimates of object's true position, orientation, and size suffice to assure convergence to a local minimum that corresponds to the actual shape. This is important since these parameters can be estimated only from the range points on the visible side of the object and hence the estimates cannot be very accurate to begin with. Initial values for both shape parameters, ϵ_1 and ϵ_2 can always be 1, which means that the initial model Λ_E is always an ellipsoid. Position in world coordinates is estimated by computing the center of gravity of all range points, and the orientation is estimated by computing the central moments with respect to the center of gravity. Estimates for model's size are simply the extent of range points along the new coordinate axis. Figure 2 shows a model recovery sequence.

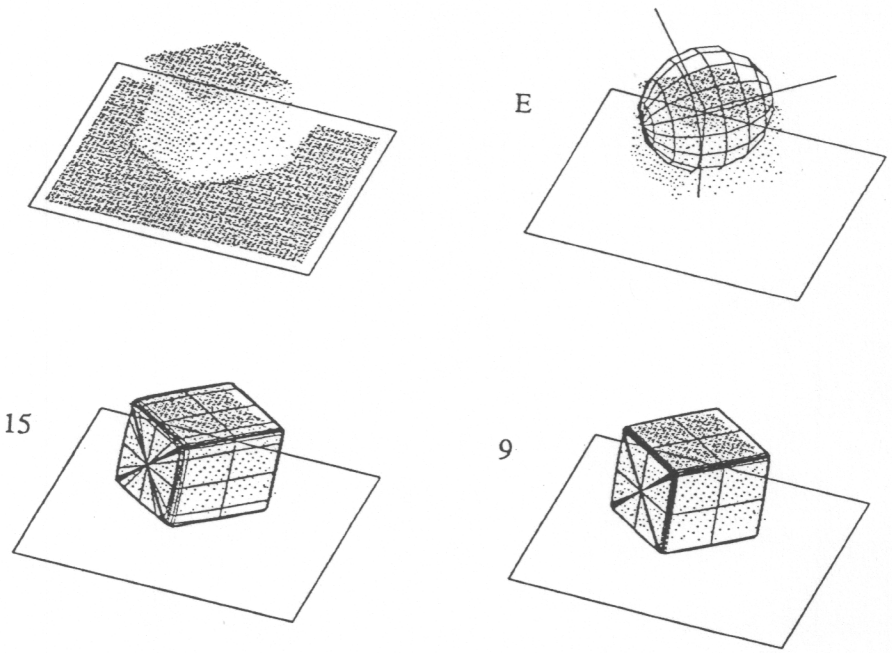

Figure 2: *Superquadric model recovery in nine iterations*

Deformed superquadrics can be recovered using the same technique as for the recovery of non-deformed superquadrics. The only difference is that some additional parameters describing deformations must also be recovered. Deformations such as simplified tapering, bending and twisting require just a few additional parameters [4]. Other types of deformations could be introduced easily. Figure 3 shows an example of deformed model recovery.

The fitting function (3) can be regarded as an energy function on the space of model parameters. Minimization methods can, in general, only guarantee convergence to a local minimum. We avoid solutions corresponding to shallow local minima by adding noise to the value of the fitting function of the accepted model at each iteration during model recovery. This stochastic technique of introducing "jitter" into the fitting procedure resembles simulated annealing.

A substantial speed up can be achieved by subsampling the original range map and using a series of coarse to fine grids during minimization since the most time consuming part of model recovery is the evaluation of the fitting function and of all of its partial derivatives for every input range point during each iteration. Thus an implementation of the recovery procedure on a fine grained parallel architecture would be straightforward since the evaluation of the fitting function and its partial derivatives is locally independent.

Recovery of models shown in this paper, where the number of range points for each model is on the order of several hundred, takes about 20 seconds of CPU time on a VAX 785 computer. For details and issues concerning consistency, stability and ambiguity of model recovery see [17].

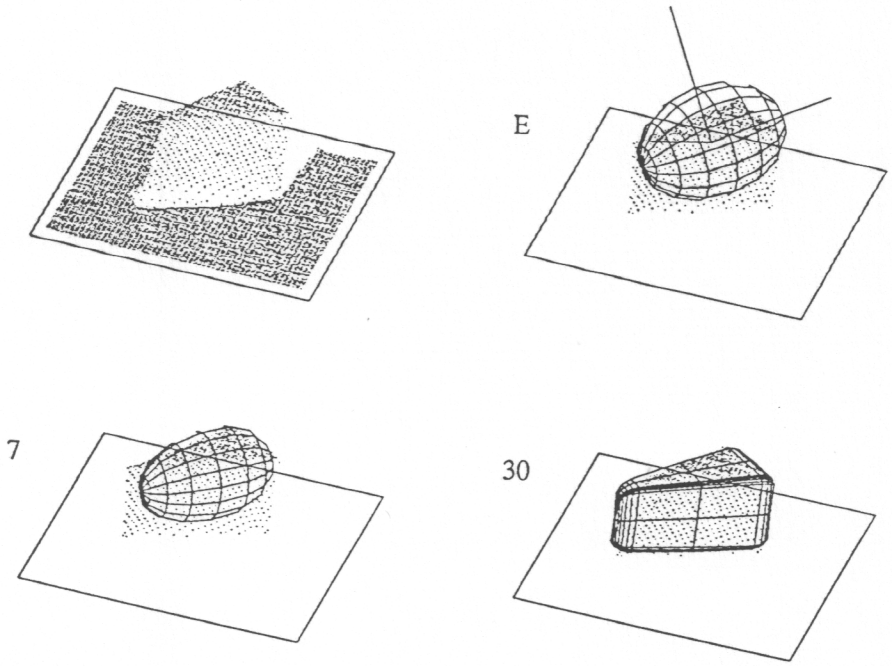

Figure 3: *Recovery of a superquadric model enhanced with tapering shape deformation (30 iterations)*

3 SEGMENTATION AS THE SEARCH FOR THE BEST DESCRIPTION IN TERMS OF PRIMITIVES

We view the segmentation process as a data reduction mechanism that requantizes the sensory measurements into some predefined primitive elements which encompass the available knowledge. This enables us to infer a symbolic description of the world. Most commonly, segmentation is viewed as a local to global aggregation problem with various similarity criteria employed to achieve a coherent global description [5]. Indeed, this global description is most usefully achieved in terms of global primitives that are easy to extract and are useful for the later processing. This can be accomplished in two ways: one is to actively use the global model as the individual primitives are being developed, in essence recovering the model as aggregation proceeds. The other way is to use a local coherence measure to first classify the data and then use the fitting technique to recover the model. The latter approach, though not limited by the global model at the aggregation stage, essentially isolates the segmentation and the representation stages, with the result that the final description might not correspond to the global model since it played no part in the segmentation process. Besides, the outliers in the data set resulting from misclassification, and the sensitivity of the methods for model estimation may lead to disastrous results [9]. As mentioned before, a desirable approach is to use both the local coherence measure and the global model to guide the segmentation, corroborating the notion that the problems of segmentation and representation are not separable [2].

In [11] we define segmentation as partitioning the images (range or reflectance) into primitive models by *searching* for the models as they are developed everywhere in the image, such that the description is best in terms of global shape and error. By searching we mean fitting and selecting only those models that best describe the underlying data using the criterion function which takes into account the number of points that are described by a particular model, its goodness-of-fit, and the structural complexity of the model. Our method performs data aggregation via model recovery in terms of variable-order (up to second-order) bi-variate patches using iterative regression. Model recovery starts simultaneously and independently at all the regions found to be globally coherent in the initial neighborhood (seed regions). All the recovered models are potential candidates for the final description. To make the method computationally feasible, it is necessary to monitor region growing and discard superfluous regions even before they are fully grown. The major novelty of this approach is in combining model extraction and model selection in a dynamic way, such that only the "best" models are allowed to develop further. Perhaps the closest in spirit to

our approach to model selection is the one used by Pentland [14]. Figures 4 and 5 show results of this segmentation process.

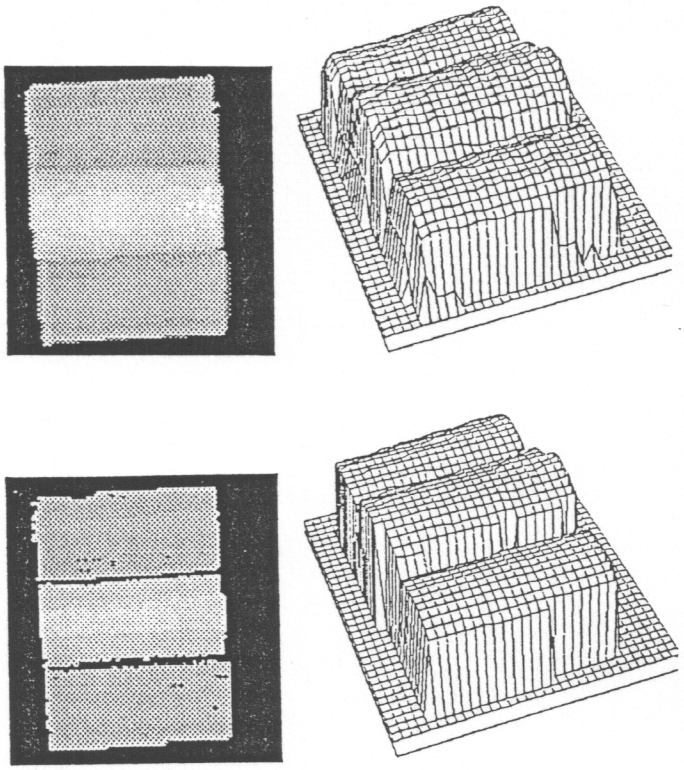

Figure 4: *The top half of the figure shows the input range data. The bottom half the results of the segmentation.*

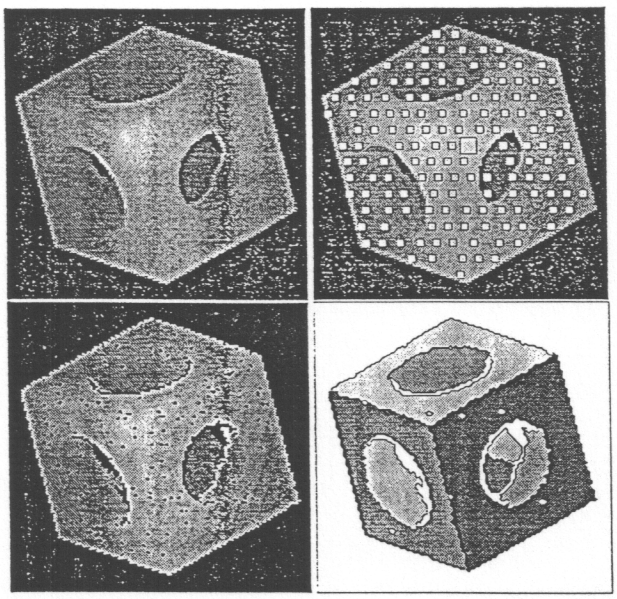

Figure 5: *The input for the segmentation algorithm are range images (top left). Segmentation starts with seed regions everywhere in the range image (top right). Only a few surface models fully developed to result in a globally coherent description while all the other regions were discarded (bottom right). Bottom left are shown only those range points that were modeled by the surface patches on the right. Dark points are noise in the original range image.*

4 CONCLUSIONS

We believe that the above segmentation schema which is presented in detail in [12] is a tool that will prove useful in many tasks of early vision. The iterative approach combining data classification and model fitting shows that segmentation and modeling are not two independent procedures but have to be integrated. The procedure which dynamically combines model recovery with model selection proves to be much more efficient than applying the modules one after another. Another important conclusion is that reliable segmentation can only be achieved by considering many competitive solutions and choosing those which reveal some kind of structure in terms of underlying models. Initial local estimates, no matter how good they are, do not necessarily lead to a good result, and more global information is needed. Optimization that is performed on the level of primitives rather than on a pixel level not only improves the performance enormously in terms of computational complexity but also gives more reliable results. The volumetric recovery scheme, on the other hand, provides a very compact shape representation that can support several tasks such as navigation, obstacle avoidance and manipulation with robot hands. We are planning to combine the above presented volumetric shape recovery and the segmentation framework in such a way to achieve the most compact possible shape representation of a given scene. When more then one shape cue is available the volumetric part recovery and segmentation method can serve as a convenient way of integrating information from different cues.

We would like to point out several possible contributions of this work to space applications. Besides the already mentioned robot manipulation [10] of rocks using superquadric models, we believe that the outlined segmentation method can be used not only for on the ground exploration by rovers (a compact symbolic description of the scene is essential to local decision making) but also during the descent phase of a landing space vehicle. Digital terrain maps provided in advance could be segmented into perceptually relevant units (peaks, valleys, planes, etc.) and represented as labeled graphs. Range data gathered during the descent could be also subjected to the same segmentation algorithm and the resulting graph compared to the stored graph that represents the complete digital terrain map of the relevant area. We believe that the comparison of the two graphs and the resulting position of the space vehicle could be performed faster than the comparison of the two actual depth maps.

5 REFERENCES

[1] P. K. Allen, K. S. Roberts, "Haptic object recognition using a multi-fingered dextrous hand," *Proceedings IEEE Conference on Robotics and Automation*, May 1989.

[2] Ruzena Bajcsy, Franc Solina, Alok Gupta: "Segmentation versus object representation–are they separable?", in R. C. Jain and A. K. Jain (Eds.), *Analysis and Interpretation of Range Images*, Springer, New York, 1990

[3] A. H. Barr: "Superquadrics and angle-preserving transformations," *IEEE Computer Graphics and Applications*, bf 1, pp. 11-23, 1981

[4] A. H. Barr: "Global and local deformations of solid primitives," *Computer Graphics*, **18**, No. 3, pp. 21-30, 1984

[5] P. J. Besl and R. Jain, "Segmentation through variable-order surface fitting," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **10**, No. 2, 1988

[6] I. Biederman, "Human image understanding: recent research and theory," *Computer Vision, Graphics, and Image Processing*, **32**, pp. 29-73, 1985

[7] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MA, 1987

[8] R. M. Bolle, B. C. Vemuri: "On Three-Dimensional Surface Reconstruction Methods," IEEE Trans. on Pattern Analysis and Machine Intelligence, **13**, No. 1, pp. 1-13, 1991

[9] D. S. Chen, "A Data-Driven Intermediate Level Feature Extraction Algorithm", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **11**, No. 7, pp. 749-758, 1989

[10] T. Choi and H. Delingette and M. DeLuise and Y. Hsin and M. Hebert and K. Ikeuchi. "A Perception and Manipulation System for Collecting Rock Samples," *Proceedings NASA Symposium on Space Operations, Applications and Research*, Albuquerque, New Mexico, 1990.

[11] A. Leonardis, A. Gupta, R. Bajcsy, "Segmentation as the Search for the Best Description of the Image in Terms of Primitives," *Proceedings 3rd IEEE International Conference on Computer Vision*, Osaka, Japan, 1990

[12] Aleš Leonardis, Alok Gupta, Ruzena Bajcsy: "Segmentation as the Search for the Best Description of the Image in Terms of Primitives", Technical report MS-CIS-90-30, GRASP Lab 215, University of Pennsylvania, 1990

[13] D. G. Lowe, T. O. Binford, *Perceptual organization and visual recognition,* Boston: Kluwer, 1985

[14] A. P. Pentland, "Automatic extraction of deformable part models," *International Journal of Computer Vision,* **4**, pp. 107-126, 1990

[15] T. Poggio, "Computational vision and regularization theory," *Nature,* **317**, 1985

[16] F. Solina, R. Bajcsy, "Recovery of mail piece shape from range images using 3-D deformable models," *International Journal of Research & Engineering, Postal Applications*, **Inaugural Issue**, pp. 125-131, 1989

[17] F. Solina, R. Bajcsy, "Recovery of parametric models from range images: The case for superquadrics with global deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, pp. 131-147, 1990